research at work

NPRST

Navy Personnel Research, Studies, and Technology
5720 Integrity Drive • Millington, Tennessee 38055-1000 • www.nprst.navy.mil

NPRST-TR-15-3 **March 2015**

Improved Forecasting Methods for Naval Manpower Studies

Ping Ying Bellamy, Ph.D.
Tanja F. Blackstone, Ph.D.
Navy Personnel Research, Studies, and Technology

NPRST-TR-15-3
March 2015

Improved Forecasting Methods for Naval Manpower Studies

Ping Ying Bellamy, Ph.D.
Tanja F. Blackstone, Ph.D.
Navy Personnel Research, Studies, and Technology

Reviewed by
Louis C. Miller, Ph.D.

Approved and released by
David M. Cashbaugh
Director

Navy Personnel Research, Studies, and Technology (NPRST)
Bureau of Naval Personnel (BUPERS-1)
5720 Integrity Drive
Millington, TN 38055-1000
www.nprst.navy.mil

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** (From-To)
25-03-2015	Technical Report	November 2012-October 2014

4. TITLE AND SUBTITLE

Improved Forecasting Methods for Naval Manpower Studies

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROJECT ELEMENT NUMBER

6. AUTHORS

Ping Ying Bellamy, Ph.D.
Tanja F. Blackstone, Ph.D.

5d. PROJECT NUMBER

P14-0149

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATIONS NAME(S) AND ADDRESS(ES)

Navy Personnel Research, Studies, & Technology (NPRST/BUPERS-1)
Bureau of Naval Personnel
5720 Integrity Drive
Millington, TN 38055-1000

8. PERFORMING ORGANIZATION REPORT NUMBER

NPRST-TR-15-3

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Forecasted manpower inventory, the number of individuals available in a given time period, are derived from stay/loss models, where estimates of the probability of staying in the navy informs the advancement and gains modules used within the Department of the Navy. As such, the accuracy of these probability rates is critical to these related functions. Extending an earlier study, we focus on two methodologies, autoregressive and logistic methods, and consider the effect of structural changes on forecast accuracy. Exogenous events or structural breaks in time-series data can result in large forecasting errors. Using the Bai-Perron (BP) test, we determine if structural breaks occur in the data. In cases where breaks are identified we control for breaks in the models. We then validate and discuss improvements in the forecast accuracy.

15. SUBJECT TERMS

Forecasting, Bai-Perron Test, continuation rates

16. SECURTIY CLASSIFICATION OF:			**17. LIMITATATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
UNCLASSIFIED			UNLIMITED	39	Wendy Douglas
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER**
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED			(901)874-2218

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Foreword

As part of a collaborative effort between Navy Personnel Research, Studies, and Technology (NRPST), and Naval Postgraduate School (NPS), the Deputy Chief of Naval Personnel, N1B, requested that a review of advances in statistical forecasting models be undertaken. The objective of the effort was to evaluate new statistical models and or methodologies that showed potential in improving the forecast accuracy of continuation rate, retention, and attrition models. The navy's manpower and personnel enterprise relies on the accuracy of forecast models to manage end-strength, skill inventories, promotions, and recruiting. Small improvements in manpower and personnel forecast models can result in significant cost reductions. Any alternative model, however, must consider the cost of model development, validation, and updating. In recognition of the costs associated with model development, the models discussed in this report focus on standard navy manpower and personnel methodologies; logistic and auto regressive. Using results from the Bai-Perron test for structural breaks, dummy variables are added to the standard models, with forecast estimates of continuation rates compared to actual continuation rates.

This effort is supported by Navy Total Force, N15. The point of contact for this effort is Dr. Tanja Blackstone, Navy Personnel Research, Studies, and Technology, (901) 874-4633.

DAVID M. CASHBAUGH

Director

Summary

Extending an earlier study (Bellamy and Blackstone 2014), this study focused on two methodologies, autoregressive and logistic methods, and considered the effect of structural changes on forecast accuracy. Exogenous events or structural breaks in time-series data can result in large forecasting errors. Using the Bai-Perron (BP) test, we first determined if structural breaks occur in the data. In cases where breaks could be identified, they were controlled for in the models.

Finally, we validated and discussed improvements in the forecast accuracy. The results show small improvements in the accuracy of forecasts for specific forecast periods when structural breaks are considered. The BP test performs better in samples with a large number of time periods. Available data used in this study only consisted of 32 quarters. To improve the accuracy of forecasts we suggest using monthly data. Using monthly data is likely to improve the overall fit of the models and the accuracy of the BP test.

A measure of unemployment to control for the effects of the economy on retention, continuation rates, and probability of loss, is generally used in navy manpower and personnel models. As this research effort was exploratory, separate models were estimated with and without unemployment variables to determine to what extent the inclusion of unemployment variable would improve forecast accuracy. The findings in this study indicate that the unemployment variable may not add to model performance but in fact worsen the overall accuracy of the model. An extension of this study would be to identify and ascertain if alternative macroeconomic variables should be considered.

Overall, the results show that controlling for structural breaks to improve the accuracy of forecasts is mixed. This study, however, only considered one skill group, the AME enlisted community and due to data limitations a small number of time periods. What is important to note is that there is no one methodology or model with consistently superior performance. Forecasting models and methodologies should be tailored to the data. Extension of this work should include use of monthly data and application of the models to other enlisted management communities.

Contents

Figures

Tables

Introduction

Loss model forecast accuracy is essential in order to assure that the Navy meets end-strength targets, efficient allocation of the manpower and personnel budgets, and to meet readiness requirements. The reliance on forecasting models to manage the Navy's manpower and personnel enterprise necessitates that analysts expend significant resources to improve forecast accuracy. These efforts entail better data quality, identifying alternative methodologies, and alternative model specifications.

Over the past decade, NPRST has investigated numerous parametric and semi-parametric methodologies to improve retention and attrition forecasts, including data mining techniques for variable selection. More recently, NPRST in collaboration with the Center for Naval Analysis (CNA) examined four methodological approaches to determine the most efficient methodology for forecasting continuation rates, (Bellamy and Blackstone, 2014). These methodologies included (1) the moving average method; (2) the pseudo-Bayesian method; (3) a combination method which uses attributes of both (1) and (2); and (4) the autoregressive method. Additionally, a logistic regression approach was used to forecast loss rates.[1] For methods (1)-(4), model specification was simplistic using current and lagged values of continuation rates. The logistic regression was slightly more complex including demographic variables, pay grade specific variables, and national unemployment.

Using absolute error differences and the difference between forecast and actual continuation rate, Bellamy and Blackstone (2014) showed marginal differences in performance in models (1)-(4). The forecast error for the moving average and autoregressive models was only marginally smaller as compared to (2) and (3). Separately, using a logistic approach to forecast the probability of staying in the Navy, improvements in forecast accuracy vary depending on pay grade and years of service.

Model performance may be affected by sample size and by structural changes not directly captured in the models (i.e., policies governing skill group contracts, promotions, bonuses, and/or exogenous events). Extending the earlier study, this study focuses on two methodologies, autoregressive and logistic methods, and considers the effect of structural changes on forecast accuracy.

Exogenous events or structural breaks in time-series data can result in large forecasting errors. Using the Bai-Perron (BP) test, we determine if structural breaks occur in the data. Secondly, in cases where breaks are identified, we control for breaks in the models. We then validate and discuss improvements in the forecast accuracy.

[1] Continuation rates are the number of separations from a group's beginning inventory during the sampling period. Separations include those individuals who have a change in PG, EMC, YOS or who are a loss to the Navy. Loss rates only capture losses to the Navy.

General Models

Autoregressive method (AR) is a statistical forecasting method that uses previous period information to forecast future periods. Lagged values of stay rates are used to forecast probability of staying, with forecast accuracy depending on the number of lags chosen. The limited number of variables provides an easy and cost effective method for predicting a given outcome variable. To estimate the probability of staying in the Navy, a simple specification of an AR model is given by:

Model (1) $SR_t = c + \sum_{i=1}^{p} \beta_i SR_{t-i} + \varepsilon_t$

where SR_t = probability of staying in the Navy (stay rate)

c = constant

β = the estimated coefficients

SR_{t-i} – lagged stay rates

p - total number of time periods (i.e. lags) considered

To control for structural changes, dummy variables are added to Model (1) giving:

Model (1a) $SR_t = c + \sum_{i=1}^{p} \beta_i SR_{t-i} + \beta_i SB_t + \varepsilon_t$

where SR_t = probability of staying in the Navy (stay rate)

c = constant

β - the estimated coefficients

SR_{t-i} – lagged stay rates

SB_t –1 if structural break, 0 otherwise

p - total number of time periods (i.e., lags) considered

The robustness of the AR methodology is dependent on the number of lag periods, 'p', included in the model. While there are statistical tests that can be used to determine the optimal number of lags, these must be applied to each period and statistically tested. As an alternative, the optimal number of lags was determined using Akaike Information Criterion (AIC); the smaller the AIC, the more robust the model.[i] For each pay grade considered in this study, separate AR models were estimated with the number of lags ranging from 1-8. Estimates of probability of stay were obtained from a baseline equation, without structural breaks, and then compared to a second equation that controlled for structural breaks. The Bai-Perron methodology was used to identify relevant structural breaks in the data.

The AR model is simple in its specification and its ease of use makes it an attractive forecasting methodology. As specified herein, the AR model is limited and does not allow for more specific analysis such as how macroeconomic factors, number of months at sea, or years of service may affect staying in the Navy. To address the limitations of Model (1), we consider a logistic model.

The general logistic model statement is given by:

Model (2) $G(z) = \exp(z)/[1 + \exp(z)] = \Lambda(z)$

The function G is between zero and one for all real number z. Model (2) estimates the probability that an individual will leave the Navy, contingent on z, where z is a matrix of explanatory variables.

Following Model (2), the estimates of the probability of leaving the Navy are obtained from Model (2a) and Model (2b) where Model (2b) includes the national unemployment rate. The national unemployment controls for employment probabilities external to the Navy that may influence the stay/leave decision.

(2a)

$$\text{Probability of Loss} = f(\text{yos seaos\_wil12m acell tig\_wi12m age cur\_compg\_mr male r\_api r\_black r\_hispanic r\_oth\_reth e1 e2 e3 e4 e6 e7})$$

(2b)

$$\text{Probability of Loss} = f(\text{yos seaos\_wil12m acell tig\_wi12m age cur\_compg\_mr male r\_api r\_black r\_hispanic r\_oth\_reth e1 e2 e3 e4 e6 e7 unemp\_r})$$

Where Loss = 1 denotes loss to the Navy; 0 otherwise.
yos = years of service
seaos_wil12m = dummy variable indicating whether the soft EAOS is coming up for the sailor in the next 12 months
acell = A cell flag
tig_wi12m = Promoted within last 12 months
age = age
cur_compg_mr = current EMC-PG inventory/ current EMC-PG BA
male = male flag. 1 denotes male; 0 denotes female
r_api otherwise = 1 if the sailor is of Asian/Pacific Islander race/ethnicity; 0
r_black = 1 if the sailor is of African American race/ethnicity; 0 otherwise
r_hispanic = 1 if the sailor is of Hispanic race/ethnicity; 0 otherwise
r_oth_reth = 1 if the sailor's race is other/unknown; 0 otherwise[2]
e1 = 1 if pay grade is e1; 0 otherwise
e2 = 1 if pay grade is e2; 0 otherwise
e3 = 1 if pay grade is e3; 0 otherwise
e4 = 1 if pay grade is e4; 0 otherwise
e6 = 1 if pay grade is e6; 0 otherwise
e7 = 1 if pay grade is e7; 0 otherwise[3]
unemp_r = national unemployment rate

[2] The benchmark for the race/ethnicity is white.
[3] The benchmark for the variable of pay grade is E5.

To obtain the probability of staying in the Navy, the probability of loss obtained from the logistics models are subtracted from '1'. Results for models (2a) and (2b), the baseline models, are provided in the following section. Based on Bai-Perron test, dummy variables were added to each of the models to control for structural breaks. Results for the logistic models which include structural breaks are compared to actual stay rates and the baseline models as specified in Model (2a) and (2b).

Description of Data used in Forecasting Loss and Stay Probabilities

For the AR model, fiscal year quarterly data from the AME enlisted community 2000-2007 was used to obtain forecast estimates for 2008-2009. Models were estimated for pay grades E4 and E6. These pay grades were chosen as, in general, they represent critical stay/leave decision points. For the AME community, individuals at E4 are at the end of their first term contract. Individuals at E6 are at mid-career or at a retirement decision point. AME quarterly data only consisted of historical stay rates; therefore, forecast estimates were based on lagged values of stay rates.

For the logistic model, annual data from 1999-2007 for the AME enlisted community was used to forecast annual stay rates by pay grade for 2008-2009. Data used in obtaining estimates for Model (2a) included individual characteristics. To control for the effect of economic conditions on stay rates, national unemployment rate was included in Model (2b).

Bai-Perron Test for Structural Breaks

The time period for the data used in this study, 1999-2009, exhibited at least two exogenous events or structural breaks that could lead to an upward or downward shift of the regression model through the constant term.[4] If the sample data is likely to exhibit breaks, it is then prudent to account for structural changes in the model to improve model performance. To test for structural changes, the standard is to apply an F-test to the entire sample and to sub-periods. Use of the F-test, however, requires a priori knowledge of the break date(s).[5] If the break points are not known with certainty, F-statistics can be obtained from a series of sub-periods and the regression with the largest F-test would indicate a possible break point. Using the same sample to identify and test for structural changes violates the assumptions of classical hypothesis testing, (Caporale and Grier, 2005).

To overcome the limitations of the F-test, Bai and Perron (1998, 2003) developed the theoretical and computational framework that allowed for multiple unknown breakpoints. The Bai-Perron methodology allows for the endogenous determination of multiple breaks by sequential comparison of restricted and unrestricted sum of squared errors, testing the null hypothesis of no structural breaks versus the alternative.

---

[4] Regime changes include 9/11/2001 and stock market downturn 2007-2008.
[5] Also known as the Chow Test (1960)

This study examines whether inclusion of structural breaks in standard military manpower and personnel models improves forecast estimates. To identify possible break points, the BP test was applied to the AME E4 and E6 pay grade data. Results are provided in Table 1 and Table 2 and are based on a critical value of .05.

Table 1
Bai-Perron Structural Change Test
AME Pay Grade E4

Bai and Perron's Multiple Structural Change Tests

supF(*l*+1 | *l*) Tests

l	supF(*l*+1 | *l*)	Pr > supF(*l*+1 | *l*)
0	23.2140144	0.0022
1	10.5051736	0.48
2	74.9742702	<.0001
3	10.5051736	0.48
4	430.439437	<.0001
5	5.65920546	0.9924

Bai and Perron's Multiple Structural Change Tests

Break Dates

Number of Breaks	Break	95% Confidence Limits	
5	2000Q2	2000Q1	2000Q3
	2001Q1	2000Q4	2001Q2
	2005Q2	2005Q1	2005Q3
	2006Q2	2005Q4	2006Q4
	2007Q1	2006Q4	2007Q2

Table 2
Bai-Perron Structural Change Test
AME Pay Grade E6

Bai and Perron's Multiple Structural Change Tests		
supF(*l*+1 | *l*) Tests		
l	supF(*l*+1 | *l*)	Pr > supF(*l*+1 | *l*)
0	4.56262824	0.9996
1	45.0724368	<.0001
2	46.6412396	<.0001
3	44.0628816	<.0001
4	44.0628816	<.0001
5	5.70692012	0.9917

Bai and Perron's Multiple Structural Change Tests			
Break Dates			
Number of Breaks	Break	95% Confidence Limits	
5	2000Q3	2000Q2	2000Q4
	2003Q1	2002Q4	2003Q2
	2003Q3	2003Q2	2003Q4
	2004Q1	2003Q4	2004Q2
	2007Q1	2006Q4	2007Q2

If Pr > supF(l+1|l), where l is the number of break points, this indicates a possible break point in the data. From Table 1, although Pr > supF(3+1|3) is greater than .05, the Pr > supF(4+1|4), is less than .05, indicating five break points. The sequential comparison process is repeated up to the point where Pr > supF(l+1|l) is greater than .05.[6]

[6] The BP test identified the break point, l, in the data. However, the actual event could have occurred in l-t. For example, a policy change that became effective in 2006 Q1, but the BP test identified the break in 2006 Q3.

The time periods identified as break points in the data along with 95% confidence levels are given in Tables 1 and 2. While the BP test provides a statistical methodology to identify the break point time periods, causes of the breaks are left to the analyst to explain. Causes of structural changes can occur in the same time period identified by the BP test or can occur with a lag. It is possible that a breakpoint identified at time l is a result of an event that occurred in l-t.

Baseline Comparisons

a. AME Enlisted Community Actual Stay Rates

All forecast estimates were compared to actual stay rates, where actual stay rate is given by equation (1).

(1) stay rate = (beginning inventory – losses)/beginning inventory

The losses and inventory counts only consider those observations that were on active duty during the previous and current periods. Individuals shown as inactive or in the reserves are excluded in the separation or inventory counts.

Actual 2008-2009 quarterly stay rates for E4 and E6 pay grades are provided in Table 3. Actual annual stay rates by pay grade for 2008-2009 are given in Table 4. For the purposes of determining improvements in forecast accuracy results obtained from the autoregressive model are compared against the baseline data in Table 3. Logistic model forecast estimates of the probability of loss are subtracted from one to obtain the probability of staying in the Navy. These are compared against the baseline data in Table 4.

Table 3
AME Actual Stay Rates
Baseline for E4 and E6 Pay Grades Autoregressive Model
Quarterly Rates 2008-2009

Paygrade	Time Period	Stay Rate	Paygrade	Time Period	Stay Rate
E4	2008Q1	0.9265	E6	2008Q1	0.9871
E4	2008Q2	0.9600	E6	2008Q2	0.9615
E4	2008Q3	0.9239	E6	2008Q3	0.9700
E4	2008Q4	0.9526	E6	2008Q4	0.9734
E4	2009Q1	0.9593	E6	2009Q1	0.9558
E4	2009Q2	0.9785	E6	2009Q2	0.9932
E4	2009Q3	0.9083	E6	2009Q3	0.9555
E4	2009Q4	0.9573	E6	2009Q4	0.9472

Table 4
AME Actual Stay Rates
Baseline for E1-E7 Pay Grades Logistic Model
Annual Rates 2008-2009

Year	Paygrade	Stay Rate
2008	E1-E3	0.883978
	E4	0.785924
	E5	0.828979
	E6	0.869565
	E7	0.793478
2009	E1-E3	0.881612
	E4	0.79351
	E5	0.875
	E6	0.871429
	E7	0.918605

Results Autoregressive Model: Baseline and with Structural Breaks

Results for the E4 AR Model (1) and Model (1a) are shown in Tables 5 and 6. Based on AIC, the dependent variable is regressed against a one period lag of stay rates. While the coefficient on the lagged variable is statistically significant for both E4 Model (1) and (1a), solely based on R-squared, overall model performance improves dramatically with the inclusion of the structural break dummy variables. Two of the structural breaks are statistically significant (Table 6); 2006 quarter 2 and 2007 quarter 1. For these time periods, there is a negative effect on the probability of staying, which could be explained by the fact that these time periods capture historical peaks in the stock and employment markets. Based on the R-squared and AIC, Model (1a) outperforms Model (1).

In comparison, the R-squared reported for Model (1) using E6 data is remarkably low. The addition of the dummy variables for the break points increases the R-squared from .0174 to .5469. R-squared should not be viewed as the sole determinate of model performance. Model (1) and (1a) include a constant term, *c*, where *c* is a non-stochastic drift parameter. In cases where time series models do not include a non-stochastic drift parameter, it is not atypical for R-squared to take on a low value. In cases where a non-stochastic drift parameter is included in the models, the usefulness of R-squared depends on the assumption that the sum of the residuals equals zero. In short, the predictive value of any model should not strictly rely on R-squared.

The coefficient on the lagged variable for the E6 regression becomes statistically significant with the inclusion of the dummy variables (see Table 7 and Table 8). The negative relationship of the previous periods stay rate to the current period may be indicators of the role of pay grade specific factors such as narrower advancement opportunities and retirement eligibility have on individual decisions to stay in the Navy. Regression results show that the third quarter for 2003 (break3) is the only statistically significant break point. Similar to E4 model results, Model (1a) performs better relative to Model (1).

Table 5
Autoregressive Baseline
AME E4 Pay Grade

Ordinary Least Squares Estimates			
SSE	0.0128544	**DFE**	29
MSE	0.0004433	**Root MSE**	0.02105
SBC	-146.58756	**AIC**	-149.46
MAE	0.0164618	**AICC**	-149.03
MAPE	1.736367	**HQC**	-148.52
Durbin-Watson	1.9265	**Regress R-square**	0.2304
		Total R-square	0.2304

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.4591	0.1681	2.73	0.0106
r_L1	1	0.5183	0.1759	2.95	0.0063

Table 6
Autoregressive with Structural Breaks
AME E4 Pay Grade

Ordinary Least Squares Estimates			
SSE	0.00731	**DFE**	24
MSE	0.0003	**Root MSE**	0.01745
SBC	-146.93	**AIC**	-156.9654
MAE	0.01269	**AICC**	-152.0958
MAPE	1.33947	**HQC**	-153.6932
Durbin-Watson	2.0679	**Regress R-square**	0.5625
		Total R-square	0.5625

Parameter Estimates					
Variable	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**
Intercept	1	0.9239	0.1981	4.66	<.0001
break1	1	-0.0281	0.0205	-1.37	0.1826
break2	1	-0.0124	0.0187	-0.66	0.5127
break3	1	-0.0281	0.0201	-1.4	0.1746
break4	1	-0.0575	0.0208	-2.77	0.0108
break5	1	-0.0515	0.0202	-2.55	0.0174
r_L1	1	0.0566	0.2099	0.27	0.7899

Table 7
Autoregressive Baseline
AME E6 Pay Grade

Ordinary Least Squares Estimates			
SSE	0.00737	**DFE**	29
MSE	0.00025	**Root MSE**	0.01594
SBC	-163.85	**AIC**	-166.71335
MAE	0.01017	**AICC**	-166.28478
MAPE	1.06094	**HQC**	-165.77846
Durbin-Watson	2.0455	**Regress R-square**	0.0174
		Total R-square	0.0174*

Parameter Estimates					
Variable	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**
Intercept	1	0.843	0.1803	4.67	<.0001
r_L1	1	0.133	0.1855	0.72	0.479

Table 8
Autoregressive with Structural Breaks
AME E6 Pay Grade

Ordinary Least Squares Estimates			
SSE	0.0034	**DFE**	24
MSE	0.00014	**Root MSE**	0.0119
SBC	-170.67	**AIC**	-180.7091
MAE	0.00847	**AICC**	-175.8395
MAPE	0.87632	**HQC**	-177.4369
Durbin-Watson	2.0837	**Regress R-square**	0.5469
		Total R-square	0.5469

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	1.3758	0.1695	8.12	<.0001
break1	1	-0.001274	0.009216	-0.14	0.8912
break2	1	-0.00806	0.0119	-0.68	0.5046
break3	1	-0.0602	0.0136	-4.43	0.0002
break4	1	-0.013	0.009209	-1.41	0.171
break5	1	-0.009401	0.0109	-0.86	0.3979
r_L1	1	-0.404	0.173	-2.33	0.0283

Results Logistic Model: Baseline and with Structural Breaks

Results for the baseline models, Model (2a) without unemployment variable, and Model (2b) controlling for unemployment, are given in Tables 9 and 10. Typically, manpower models include one or more variables to control for the effect of economic conditions on probability of staying (or loss) to the Navy. In general, the national unemployment rate is used. As this research effort was exploratory, separate models were estimated with and without an unemployment variable to determine to what extent the inclusion of unemployment variable would improve forecast accuracy.

Comparison of AIC between Model (2a) and Model (2b) indicate little difference in the performance of the models. The parameter estimate for unemployment in Model (2b) is statistically significant; however, the sign on the coefficient is positive. Earlier work done on retention models (Golan and Blackstone, 2008) not only reported a negative relationship between unemployment and retention for some skill groups, but their findings also showed a decreasing sensitivity of changes in unemployment on retention decisions. As noted in Golan and Blackstone (2008), there are probably a number of explanations for this result, including changes in personnel quality and the Navy's rules and policies governing the management of personnel. The effect of unemployment on the loss probability is very small and viewed in isolation does not appear to improve forecast estimates.

Dummy variables to control for structural breaks were then included in Models (2a) and (2b). Results for the models which allow for structural breaks in the data are given in Tables 11 and 12. Data to estimate the coefficients for the logistic model used annual data as such the number of time periods was insufficient to allow application of the BP test to identify structural breaks. As a proxy, the break points in the logistic model were based on the 95% confidence limits for the Bai-Perron test applied to the E4 AME data. See Table 1. Break point dates included in the logistic models are 2000, 2001, 2005, 2006, and 2007.

AIC is slightly smaller when structural breaks are included in the models. The statistical significance of the break points differs across models. In Model (2a), 2005 is the only statistically significant break point in contrast to 2001 in Model (2b). Several factors could explain this incongruent result. In 2001, unemployment increased by 1.5 percentage points, from 4.2 (January 2001) to 5.7 percent (December 2001). The significance of the 2001 structural break may account for the sudden rise (shift) in unemployment for this period. However, one would expect to observe this same effect in Model (2a). A more likely explanation is that the proxy break points derived from the quarterly data are not conducive to the annual data.

Table 9
Logistic Baseline without Unemployment
Model 2a

Analysis of Maximum Likelihood Estimates					Model Fit Statistics	
Parameter	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**	**Criterion**	**Intercept and Covariates**
Intercept	-4.2529	0.3421	154.5195	<.0001	**AIC**	8325.791
yos	0.0848	0.0145	34.422	<.0001	**SC**	8461.846
seaos_wil12m	2.7541	0.068	1639.556	<.0001	**-2 Log L**	8289.791
acell	-0.0679	0.0605	1.2601	0.2616		
tig_wi12m	-0.0463	0.072	0.4144	0.5198		
age	0.0171	0.00933	3.3526	0.0671		
cur_compg_mr	0.2382	0.1862	1.6368	0.2008		
male	-0.4385	0.1167	14.1061	0.0002		
r_api	-0.2633	0.1741	2.2862	0.1305		
r_black	0.0577	0.1129	0.2612	0.6093		
r_hispanic	-0.3271	0.141	5.3831	0.0203		
r_oth_reth	0.1095	0.1714	0.4082	0.5229		
e1	2.3669	0.1908	153.8141	<.0001		
e2	1.9174	0.1542	154.6079	<.0001		
e3	1.5006	0.1196	157.4829	<.0001		
e4	0.986	0.0966	104.205	<.0001		
e6	-1.0601	0.1235	73.7098	<.0001		
e7	-1.0497	0.176	35.5621	<.0001		

Table 10
Logistic Baseline with Unemployment
Model 2b

Analysis of Maximum Likelihood Estimates					Model Fit Statistics	
Parameter	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**	**Criterion**	**Intercept and Covariates**
Intercept	-4.7805	0.3784	159.5617	<.0001	**AIC**	8316.847
yos	0.0871	0.0145	36.0223	<.0001	**SC**	8460.461
seaos_wil12m	2.7775	0.0686	1639.34	<.0001	**-2 Log L**	8278.847
acell	-0.0602	0.0606	0.9881	0.3202		
tig_wi12m	-0.0372	0.0721	0.2661	0.606		
age	0.0164	0.00933	3.0926	0.0786		
cur_compg_mr	-0.058	0.2064	0.0789	0.7788		
male	-0.4493	0.1167	14.8137	0.0001		
r_api	-0.2655	0.1742	2.3225	0.1275		
r_black	0.056	0.113	0.2456	0.6202		
r_hispanic	-0.3322	0.1411	5.5437	0.0185		
r_oth_reth	0.1054	0.1716	0.3768	0.5393		
e1	2.4311	0.1922	160.0474	<.0001		
e2	1.9625	0.1549	160.4883	<.0001		
e3	1.5608	0.1209	166.7457	<.0001		
e4	0.9926	0.0966	105.6074	<.0001		
e6	-1.0603	0.1241	73.0484	<.0001		
e7	-1.088	0.1773	37.667	<.0001		
unemp_r	0.1609	0.0486	10.971	0.0009		

Table 11
Logistic Baseline without Unemployment
Structural Breaks Model 2a

Analysis of Maximum Likelihood Estimates				
Parameter	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	-4.1145	0.3507	137.6079	<.0001
break1	-0.1002	0.1262	0.6304	0.4272
break2	0.1296	0.104	1.5529	0.2127
break3	0.4083	0.1296	9.9346	0.0016
break4	0.1621	0.1244	1.6972	0.1927
break5	0.2135	0.1246	2.9368	0.0866
yos	0.0896	0.0146	37.9215	<.0001
seaos_wil12m	2.7583	0.0684	1626.681	<.0001
acell	-0.078	0.0608	1.6435	0.1999
tig_wi12m	0.00491	0.0736	0.0044	0.9469
age	0.016	0.00936	2.9301	0.0869
cur_compg_mr	-0.0321	0.2059	0.0243	0.8761
male	-0.4609	0.1164	15.6676	<.0001
r_api	-0.2816	0.1744	2.6065	0.1064
r_black	0.0487	0.113	0.186	0.6663
r_hispanic	-0.3595	0.1414	6.4675	0.011
r_oth_reth	0.0822	0.1718	0.2289	0.6323
e1	2.3926	0.1915	156.0905	<.0001
e2	1.9537	0.1548	159.3365	<.0001
e3	1.5204	0.1205	159.1987	<.0001
e4	0.9765	0.0965	102.3335	<.0001
e6	1.0686	0.1237	74.6787	<.0001
e7	1.1084	0.1775	38.9961	<.0001

Model Fit Statistics		
Criterion	**Intercept**	**Intercept and Covariates**
AIC	10743.6	8317.434
SC	10751.15	8491.282
-2 Log L	10741.595	8271.434

Table 12
Logistic Baseline with Unemployment
Structural Breaks Model 2b

Analysis of Maximum Likelihood Estimates

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	-5.6312	0.5135	120.2553	<.0001
break1	0.000831	0.1286	0	0.9948
break2	-0.363	0.1613	5.0611	0.0245
break3	0.1177	0.1481	0.6317	0.4267
break4	0.0322	0.1285	0.0628	0.8021
break5	0.07	0.1295	0.2925	0.5886
yos	0.0906	0.0146	38.6659	<.0001
seaos_wil 12m	2.7704	0.0686	1630.067	<.0001
acell	-0.0784	0.0609	1.6588	0.1978
tig_wi12m	0.0165	0.0738	0.0503	0.8226
age	0.0159	0.00937	2.8809	0.0896
cur_comp g_mr	-0.2618	0.2119	1.5263	0.2167
male	-0.4629	0.1164	15.8254	<.0001
r_api	-0.287	0.1746	2.7005	0.1003
r_black	0.0474	0.1132	0.1756	0.6752
r_hispanic	-0.3653	0.1416	6.6592	0.0099
r_oth_reth	0.0765	0.1721	0.1977	0.6566
e1	2.4458	0.1926	161.2949	<.0001
e2	1.9857	0.1554	163.3444	<.0001
e3	1.5576	0.1209	165.8983	<.0001
e4	0.9779	0.0966	102.4997	<.0001
e6	-1.0616	0.124	73.2465	<.0001
e7	-1.1308	0.1782	40.2504	<.0001
	0.4012	0.0984	16.6375	<.0001

Model Fit Statistics

Criterion	Intercept	Intercept and Covariates
AIC	10743.6	8302.331
SC	10751.15	8483.737
-2 Log L	10741.595	8254.331

Forecast Accuracy: Cross Model Comparison

Data from 1999-2007 were used to obtain coefficient estimates for the AR (1) model and logistic models. The estimates were then used to forecast the probability of staying for 2008 and 2009. For the AR (1) model forecast estimates are given by quarter with yearly forecast estimates calculated for the logistic regression. Visual comparisons of model performance are given in figures 1-6.

Figures 1 and 2 illustrate the differences in Model (1) and Model (1a) forecast estimates with and without structural breaks as compared to actual E-4 pay grade stay rates. With the exception of quarter three for both 2008 and 2009, inclusion of structural breaks in the AR (1) model exhibits slightly poorer performance relative to the baseline. This same result holds for the E6 pay grade, (see Figure 3 and Figure 4).

The estimates for the AR (1) models were obtained from a limited number of observations; 32 quarters. In general, AR models perform better with a large number of observations. Forecast accuracy would be improved with monthly observations, however, these more frequent observations were unavailable for this study.

Comparison of actual stay rates and forecast estimates obtained from the logistic model are given in Figures 5 and 6. For purposes of consistency, Figures 5 and 6 only provide the forecast comparisons for E4 and E6 pay grades. For the 2008 E4 pay grade, the Model 2a and 2b baseline performs slightly better relative to the 2a and 2b models with breaks included. In contrast, the E6 forecast estimates obtained from Model (2b) with breaks, shows a marginal improvement in forecast accuracy. For 2008 and 2009, Model (2a) baseline and with breaks show nearly identical forecast estimates, outperforming both logistic models that include the unemployment variable.

The model comparisons show that no one model stands out as a consistently best forecast model. The study results indicate that accuracy of the forecast varies by methodology, model, time period, and pay grade. In all cases model performance was poorer for those models that included unemployment as a macroeconomic control variable. This statement is caveated to the skill group, time period, and methodology used in this study and should not be generalized to other methodologies, data, or models.

Figure 1
Model 1 Comparison of Actual and Forecasted Stay Rate
Pay Grade E4
2008 Quarterly

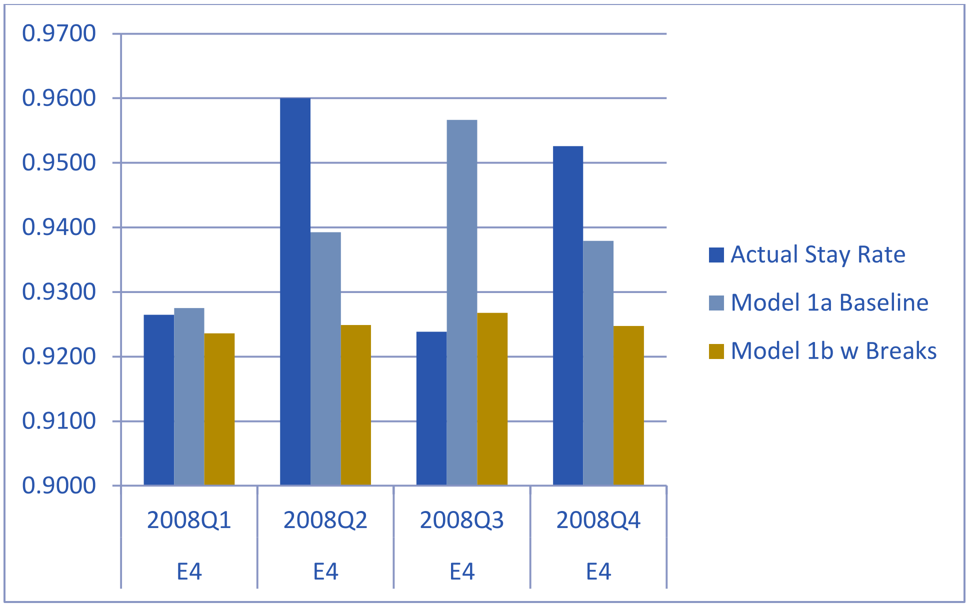


Figure 2
Model 1 Comparison of Actual and Forecasted Stay Rate
Pay Grade E4
2009 Quarterly

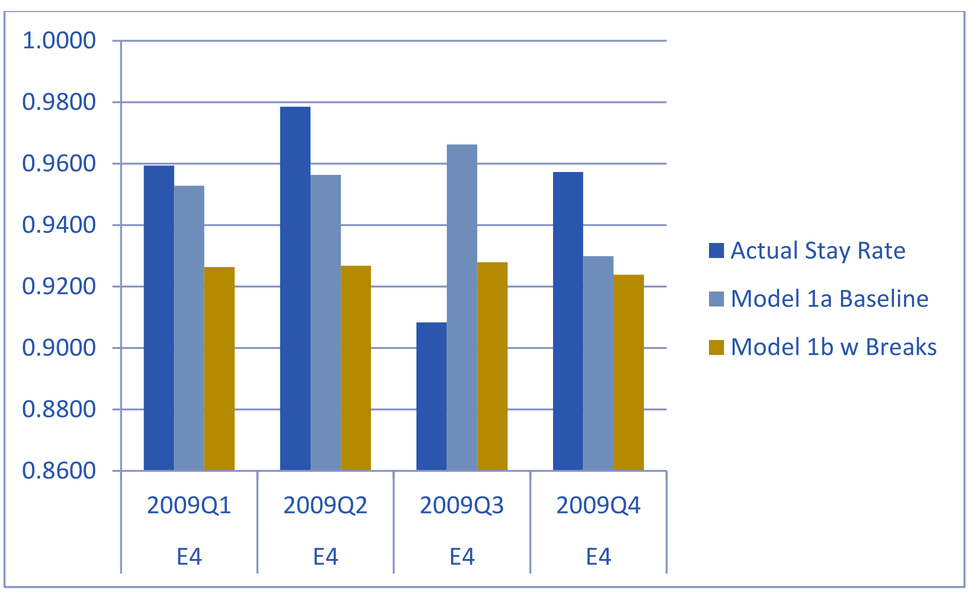


Figure 3
Model 1 Comparison of Actual and Forecasted Stay Rate
Pay Grade E6
2008 Quarterly

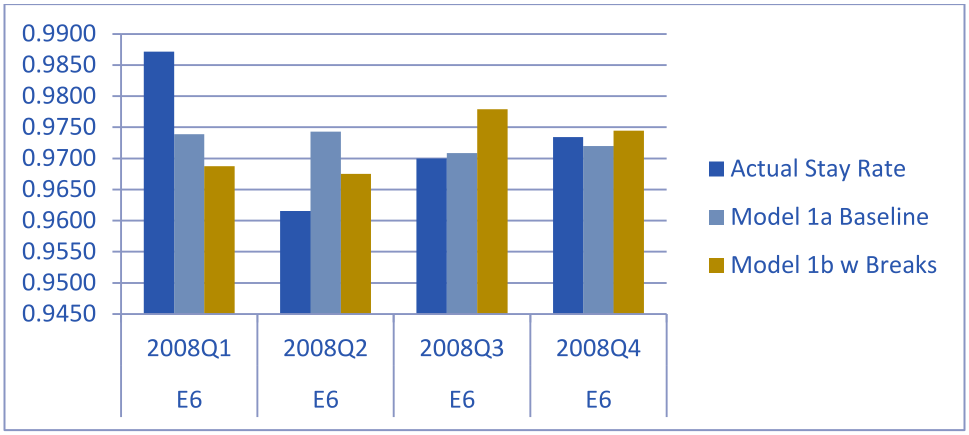


Figure 4
Model 1 Comparison of Actual and Forecasted Stay Rate
Pay Grade E6
2009 Quarterly

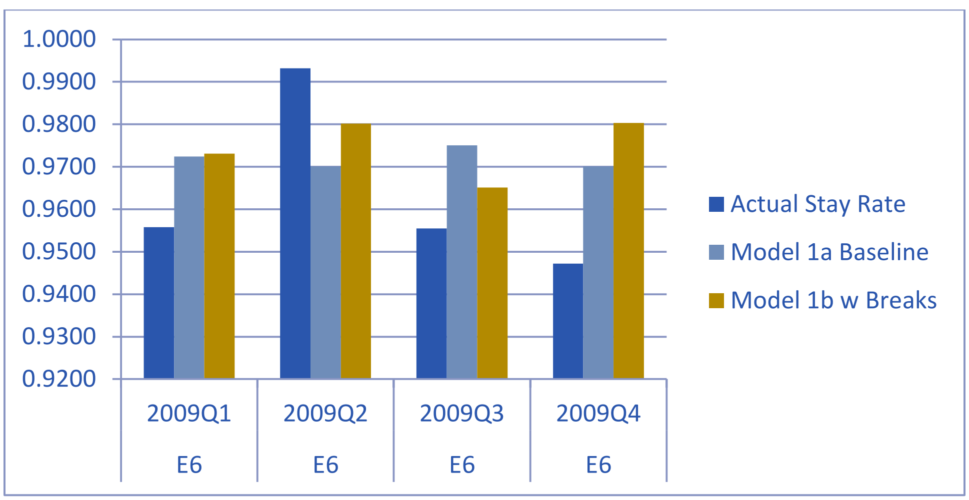


Figure 5
Model 2 Comparison of Actual and Forecasted Stay Rate
Pay Grade E4 and E6
2008

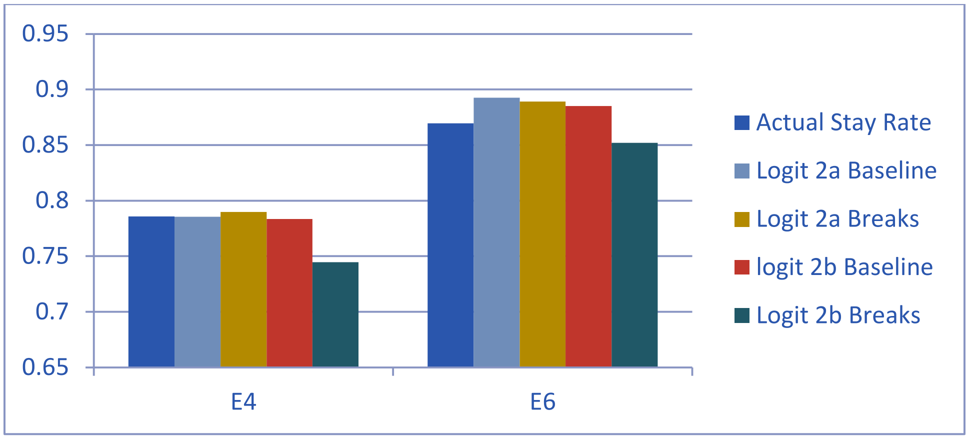
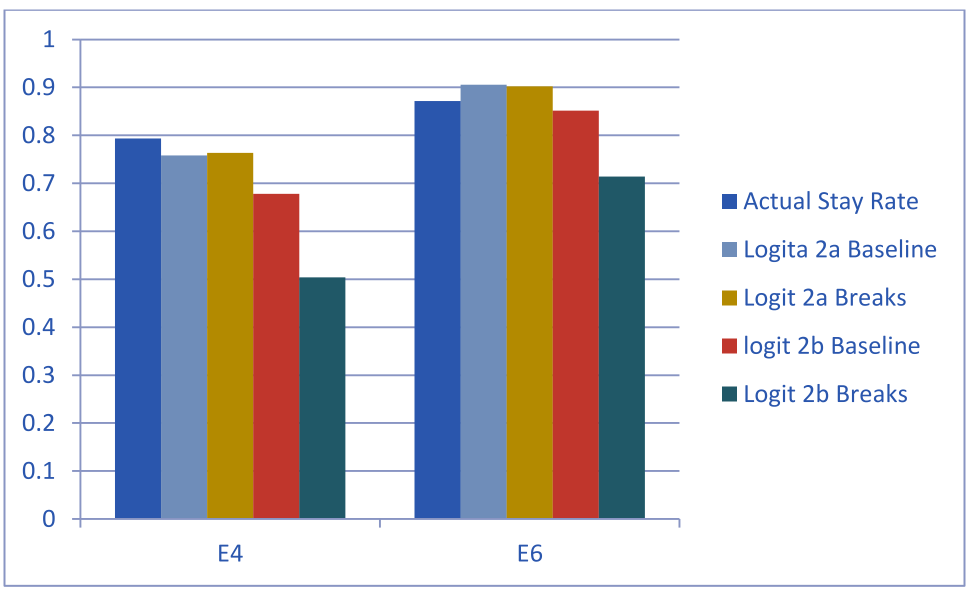


Figure 6
Model 2 Comparison of Actual and Forecasted Stay Rate
Pay Grade E4 and E6
2009



Conclusion

Extending an earlier study (Bellamy and Blackstone 2014), this study focused on two methodologies, autoregressive and logistic methods, and considered the effect of structural changes on forecast accuracy. Exogenous events or structural breaks in time-series data can result in large forecasting errors. Using the Bai-Perron (BP) test, we first determined if structural breaks occur in the data. In cases where breaks could be identified, they were controlled for in the models.

Finally, we validated and discussed improvements in the forecast accuracy. The results show small improvements in the accuracy of forecasts for specific forecast periods when structural breaks are considered. The BP test performs better in samples with a large number of time periods. Available data used in this study only consisted of 32 quarters. To improve the accuracy of forecasts we suggest using monthly data. Using monthly data is likely to improve the overall fit of the models and the accuracy of the BP test.

A measure of unemployment to control for the effects of the economy on retention, continuation rates, and probability of loss, is generally used in navy manpower and personnel models. As this research effort was exploratory, separate models were estimated with and without unemployment variables to determine to what extent the inclusion of unemployment variable would improve forecast accuracy. The findings in this study indicate that the unemployment variable may not add to model performance but in fact worsen the overall accuracy of the model. An extension of this study would be to identify and ascertain if alternative macroeconomic variables should be considered.

Overall, the results show that controlling for structural breaks to improve the accuracy of forecasts is mixed. This study, however, only considered one skill group, the AME enlisted community and due to data limitations a small number of time periods. What is important to note is that there is no one methodology or model with consistently superior performance. Forecasting models and methodologies should be tailored to the data. Extension of this work should include use of monthly data and application of the models to other enlisted management communities.

References

[1] *Forecasting Navy Continuation Rates: An Exploratory Analysis*, Yevgeniya K. Pinelis with Michael J. Moskowitz, David M. Rodney, Edward J. Schmitz, Warren T. Sutton, David L. Reese, Jessica S. Oi, Robert W. Shuford, January 2011.

[2] *The Design of the Execution Monitoring Models for Enlisted and Officer Community Management*, Edward J. Schmitz • Amanda Kraus • Warren T. Sutton Yevgeniya K. Pinelis, July 2012.

[3] *IMPACT-AC: Pseudo-Bayesian Loss Model Specification*, Icosystem, May 4, 2011.

[4] *IMPACT-AC: Pseudo-Bayesian Loss Model Specification*, Icosystem, August 2, 2012.

[5] David M. Rodney. *A Pseudo-Bayesian Technique for Estimating Continuance in PERMIT, a Personnel Management Integration Tool,* Feb 1991 (CNA Research Memorandum 90-208).

[6] Ping Ying Bellamy and Tanja F. Blackstone. *Validation of Loss and Continuation Rate Models to Support Navy Community Management*, January 2014, NPRST-TR-14-2.

[7] Yvonne M. Bishop, Stephen E. Fienberg, and Paul W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. New York: Springer, 2007.

[8] Bai, J., Perron P. 1998, Estimating and Testing Linear Models with Multiple Structural Changes, Econometrica 66:47-78.

[9] Bai J., Perron, P., 2003, Computation an Analysis of Multiple Structural Change Models, Journal of Applied Econometrics 18: 1-22.

[10] Caporale T., Grier K. 2005, How Smart is my Dummy? Time Series Test for the Influence of Politics, Political Analysis 13:77-94.

[11] Golan, A, Blackstone, T., Models of Compensation (MODCOMP): Policy Analyses and Unemployment Effects, NPRST-TR-08-3, August 2008.

Appendix A

[1] From http://en.wikipedia.org/wiki/Akaike_information_criterion. The Akaike information criterion is a measure of the relative goodness of fit of a statistical model. It is grounded in the concept of information entropy, in effect, offering a relative measure of the information lost when a given model is used to describe reality. It can be said that it describes the trade-off between bias and variance in model construction, or loosely speaking, between accuracy and complexity of the model.

AIC values provide a means for model selection. In the general case, the AIC is

$$AIC = 2k - 2\ln(L)$$

where k is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model.

Given a set of candidate models for the data, *the preferred model is the one with the minimum AIC value.* Hence AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages overfitting (increasing the number of free parameters in the model improves the goodness of the fit, regardless of the number of free parameters in the data-generating process).

AIC is founded in information theory. Suppose that the data is generated by some unknown process f. We consider two candidate models to represent f: g_1 and g_2. If we knew f, then we could find the information lost from using g_1 to represent f by calculating the Kullback–Leibler divergence, $D_{KL}(f,g_1)$; similarly, the information lost from using g_2 to represent f would be found by calculating $D_{KL}(f,g_2)$. We would then choose the candidate model that minimized the information loss.

We cannot choose with certainty, because we do not know f. Akaike (1974) showed, however, that we can estimate, via AIC, how much more (or less) information is lost by g_1 than by g_2. It is remarkable that such a simple formula for AIC results. The estimate, though, is only valid asymptotically; if the number of data points is small, then some correction is often necessary (see AICc, below).

How to apply AIC in practice: To apply AIC in practice, we start with a set of candidate models and then find the models' corresponding AIC values. There will almost always be information lost due to using one of the candidate models to represent the "true" model. We wish to select, from among R candidate models, the model that minimizes the information loss. We cannot choose with certainty, but we can minimize the estimated information loss.

Denote the AIC values of the candidate models by AIC_1, AIC_2, AIC_3, …, AIC_R. Let $\mathrm{AIC}_{\min}$ be the minimum of those values. Then $\exp((\mathrm{AIC}_{\min}-\mathrm{AIC}_i)/2)$ can be interpreted as the relative probability that the ith model minimizes the (estimated) information loss.[2]

As an example, suppose that there were three models in the candidate set with AIC values 100, 102, and 110. Then the second model is $\exp((100-102)/2) = 0.368$ times as probable as the first model to minimize the information loss, and the third model is $\exp((100-110)/2) = 0.007$ times as probable as the first model to minimize the information loss. In this case, we would

omit the third model from further consideration. We could take a weighted average of the first two models, with weights 1 and 0.368, respectively, and then do statistical inference based on the weighted multi-model;[3] alternatively, we could gather more data to distinguish between the first two models.

If all the models in the candidate set have the same number of parameters, then using AIC might at first appear to be very similar to using the likelihood-ratio test. There are, however, important distinctions. In particular, the likelihood-ratio test is valid only for nested models whereas AIC (and AICc) has no such restriction.[4]

The quantity $\exp((\mathrm{AIC}_{\min}-\mathrm{AIC}_i)/2)$ is the relative likelihood of model i.

Distribution

AIR UNIVERSITY LIBRARY
ARMY RESEARCH INSTITUTE LIBRARY
ARMY WAR COLLEGE LIBRARY
CENTER FOR NAVAL ANALYSES LIBRARY
HUMAN RESOURCES DIRECTORATE TECHNICAL LIBRARY
JOINT FORCES STAFF COLLEGE LIBRARY
MARINE CORPS UNIVERSITY LIBRARIES
NATIONAL DEFENSE UNIVERSITY LIBRARY
NAVAL HEALTH RESEARCH CENTER WILKINS BIOMEDICAL LIBRARY
NAVAL POSTGRADUATE SCHOOL DUDLEY KNOX LIBRARY
NAVAL RESEARCH LABORATORY RUTH HOOKER RESEARCH LIBRARY
NAVAL WAR COLLEGE LIBRARY
NPRST SPISHOCK LIBRARY
OFFICE OF NAVAL RESEARCH (CODE 34)
PENTAGON LIBRARY
USAIR FORCE ACADEMY LIBRARY
US COAST GUARD ACADEMY LIBRARY
US MERCHANT MARINE ACADEMY BLAND LIBRARY
US MILITARY ACADEMY AT WEST POINT LIBRARY
US NAVAL ACADEMY NIMITZ LIBRARY